

Guest Editorial
**Special Issue: Current Methodological
Issues in Educational Large-Scale
Assessments – Part II**

Matthias Stadler¹, Samuel Greiff² & Sabine Krolak-Schwerdt²

This is the second part of the special issue on current methodological issues in educational large-scale assessments (LSAs). The first part of the special issue, which was published at the end of last year (Stadler, Greiff, & Krolak-Schwerdt, 2016), consisted of four papers highlighting the diversity of challenges currently faced within large-scale assessments (LSA) while simultaneously introducing potential solutions. The authors, all established experts in the field of LSA, demonstrated exciting new ways of handling the transition to computer-based testing, maintaining maximum measurement precision, and dealing with missing data.

This second part of our special issue will now primarily focus on the handling of latent variables in LSAs through advanced item response theory (IRT) based approaches, which require sample sizes that are not commonly found in smaller sized studies. Most LSAs allow for approaches to data analysis that would not be feasible in smaller samples and enable the investigation of specific research questions regarding the nature of the underlying data. In addition, LSAs are currently going through a transition from paper-based to computer-based test administration (Bürger, Kröhne, & Goldhammer, 2016) that provides the opportunity to record not only the response of the examinee but many other variables including the item-specific response time, or behavioral patterns displayed during the assessment. Obviously, such research produces even larger amounts of data and calls for the development of new methods that are suitable to deal with this type of data.

The second part of our special issue will, therefore, include five more papers that deal with advanced approaches of tackling specific problems that researchers encounter when planning, conducting, and interpreting the results of educational large-scale assessments.

¹ *Correspondence concerning this article should be addressed to:* Matthias Stadler, PhD, ECCS unit, University of Luxembourg, 11 Porte des Sciences, 4366 Esch-Sur-Alzette, Luxembourg; email: matthias.stadler@uni.lu

² ECCS unit, University of Luxembourg, Luxembourg

Next to introducing their latest research results, the authors of all papers included valuable guidance for other researchers intending to conduct their own research in the exciting field of educational LSAs.

In the first paper of this second part of the special issue, entitled “Large-scale assessments: potentials and challenges in longitudinal designs“, Jutta von Maurice, Sabine Zinn, and Ilka Wolter elaborated on the benefits and challenges of implementing a longitudinal design into large-scale assessments in educational research. The focus of this paper is on educational trajectories and competence development. In this, the paper functions as a guideline for conducting a longitudinal large-scale study as well as for the maintenance of a panel study. Based on the starting cohort of ninth graders of the German National Educational Panel Study as an example, the authors provided detailed information on the aspects of sampling and selectivity, instrumentation, challenges and methods connected with the measurement of competence in a longitudinal large-scale study.

The second paper by Leslie Rutkowski entitled “Design considerations for planned missing auxiliary data in a latent regression context” deals with missing background data in educational LSAs and discusses possible questionnaire designs that create an improved foundation from which missing background questionnaire data can be imputed. Among these design features, splitting constructs across questionnaires, planning for missing data among well-correlated constructs, and administering intensive questionnaires to a smaller subsample are considered as potential solutions. In addition to the feasibility of each design, gains in information and the multidimensional burden of preserving achievement distributions are considered.

The third paper “Recent IRT approaches to test and correct for response styles in PISA background questionnaire data: a feasibility study” by Lale Khorramdel, Matthias von Davier, Jonas P. Bertling, Richard D. Roberts, and Patrick C. Kyllonen introduces a new IRT approach and its multidimensional extension to test and correct for response styles within large-scale assessments. Based on data from the Programme for International Student Assessment 2012 field trial, the aim of the study is to look at the potential of this new methodology to test for and correct response style bias in an international context. The authors demonstrate that personality scales corrected for response styles can lead to more valid test scores and may even address the “paradoxical relationship” phenomenon of negative correlations between personality scales and cognitive proficiencies found in various educational LSAs (Van de Gaer, Grisay, Schulz, & Gebhardt, 2012).

Christine Hohensinn and Klaus D. Kubinger contributed the fourth paper entitled “Using Rasch model generalizations for taking testee’s speed, in addition to their power, into account”. The paper deals with the concept of awarding quick solutions in ability tests with bonus points in order to gain more information about a testee’s ability. In order to determine whether speed and power do actually measure the same ability, the authors compared Rasch’s multi-dimensional polytomous model as well as his unidimensional polytomous model and Fischer’s speed-and-power two-steps model. Their results indicate that speed and power are in fact different dimensions of the same construct rather than largely independent dimensions.

Finally, Zhan Shu, Yoav Bergner, Mengxiao Zhu, Jiangang Hao, and Alina A. von Davier introduced their approach to characterize and capture the unique features of each individual's response process during a problem-solving activity in scenario-based tasks in the paper entitled "An Item Response Theory analysis of problem-solving processes in scenario-based tasks". The introduction of computer-based assessment in educational LSAs has also brought the option to analyze students' behaviour when solving tasks. The authors introduced the structure, assumptions, parameter space, and the estimation of the parameters of a Markov-IRT model designed to analyze such process data. They further discuss its usefulness in characterizing students' response processes using an empirical example based on a scenario-based task from the National Assessment of Educational Progress – Technology & Engineering Literacy Assessment (NAEP–TEL).

We hope that readers will also enjoy this second part of our special issue and find it helpful for their own research. Moreover, we are particularly grateful to Klaus Kubinger, the Editor-in-Chief of the journal for hosting this special issue in *Psychological Test and Assessment Modeling*.

References

- Bürger, S., Kröhne, U., & Goldhammer, F. (2016). The transition to computer-based testing in large-scale assessments: Investigating (partial) measurement invariance between modes. *Psychological Test and Assessment Modeling*, *58*, 587-606.
- Stadler, M., Greiff, S., & Krolak-Schwerdt, S. (2016). Current methodological issues in educational large-scale assessments. Guest editorial. *Psychological Test and Assessment Modeling*, *58*, 593-595.
- Van de Gaer, E., Grisay, A., Schulz, W., & Gebhardt, E. (2012). The reference group effect. An explanation of the paradoxical relationship between academic achievement and self-confidence across countries. *Journal of Cross-Cultural Psychology*, *43*, 1205-1228.